THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

University-level practical activities in bioinformatics benefit voluntary groups of pupils in the last 2 years of school

OPEN ACCESS

International Journal of
STEM Education
a SpringerOpen Journal

SHORT REPORT   **Open Access**

CrossMark

University-level practical activities in bioinformatics benefit voluntary groups of pupils in the last 2 years of school

Daniel Barker[1*], Rosanna G. Alderson[2], James L. McDonagh[2,3], Heleen Plaisier[1], Muriel M. Comrie[4,5], Leigh Duncan[6], Gavin T. P. Muirhead[4] and Stuart D. Sweeney[6,7]

Abstract

Background: Bioinformatics—the use of computers in biology—is of major and increasing importance to biological sciences and medicine. We conducted a preliminary investigation of the value of bringing practical, university-level bioinformatics education to the school level. We conducted voluntary activities for pupils at two schools in Scotland (years S5 and S6; pupils aged 15–17). We used material originally developed for an optional final-year undergraduate module and now incorporated into 4273π, a resource for teaching and learning bioinformatics on the low-cost Raspberry Pi computer.

Results: Pupils' feedback forms suggested our activities were beneficial. During the course of the activity, they provide strong evidence of increase in the following: pupils' perception of the value of computers within biology; their knowledge of the Linux operating system and the Raspberry Pi; their willingness to use computers rather than phones or tablets; their ability to program a computer and their ability to analyse DNA sequences with a computer. We found no strong evidence of negative effects.

Conclusions: Our preliminary study supports the feasibility of bringing university-level, practical bioinformatics activities to school pupils.

Keywords: Bioinformatics, Computational biology, Secondary school, Raspberry Pi, Open access teaching material, Case study

Findings

Introduction

Progress in Science, Technology, Engineering, Mathematics and Medicine (STEMM) subjects is increasingly dominated by computational analyses. In biological sciences, for example, the exceptional pace of recent advances in technology for DNA and genome sequencing has created a demand for computationally able researchers, to analyse the large amounts of data produced. A field specialising in application of computation to biological problems has emerged, known as bioinformatics. The development of bioinformatics is discussed by Hogeweg (2011), and university-level bioinformatics education has been reviewed by Magana et al. (2014).

DNA sequences and related data are available at low cost (for new sequencing work) or free in online databases such as GenBank (Benson et al. 2015), Ensembl (Cunningham et al. 2015) and hundreds of others (Galperin et al. 2015). Software for bioinformatics research is usually free, for example the very widely used sequence database search software, BLAST (Altschul et al. 1997). Free resources are also available for bioinformatics teaching and learning, for example 4273π (Barker et al. 2013), Bioinformática na escola (Marques et al. 2014), GOBLET (Corpas et al. 2015), Bioinformatics@school (http://www.nbic.nl/nl/education/high-school-programmes/bioinformaticsschool) and the EvoEd Digital Library (http://evoed.evolutionsociety.org). These publicly available data, software and materials present excellent opportunities for relatively low-cost teaching.

* Correspondence: db60@st-andrews.ac.uk
[1]Sir Harold Mitchell Building, School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, UK
Full list of author information is available at the end of the article

Springer

Barker *et al. International Journal of STEM Education* (2015) 2:17

Page 2 of 8

There has been a recent, encouraging increase in exposure of school pupils to bioinformatics (e.g. Gallagher et al. 2011; Lewitter and Bourne 2011; McQueen et al. 2012; Kovarik et al. 2013; Machluf and Yarden 2013; Wood and Gebhardt 2013; Marques et al. 2014; Toby and Pope 2014). Genomics and associated topics have started to appear in many official school curricula, for example in Scotland (see "Discussion", below), the Netherlands (College voor Examens 2014, p. 17) and the USA (Wefer and Sheppard 2008). From a different angle, computer science is now a major part of the primary school curriculum for England (https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study). This is in line with a "back to basics" approach to computing currently emerging, as opposed to more traditional information and communications technology (ICT). In the UK, this change has been particularly associated with the low-cost Raspberry Pi computer, which is suitable for educational projects in electronics and engineering as well as general use and has sold over 5 million units (http://www.raspberry-pi.org; http://www.wired.co.uk/news/archive/2015-02/18/raspberry-pi-5-million). However, a practical link between computers and STEMM—which we will refer to as computational science, as opposed to computer science—still does not feature strongly on the UK school curriculum. DNA sequencing has a pervasive and increasing influence across traditionally disparate subject areas, including biochemistry, biomedical research, clinical medicine, evolutionary biology, ecology, neuroscience and anthropology. DNA sequencing is used to diagnose genetic and infectious diseases, discover drugs, characterise environments, monitor the progress of cancers, identify species and reveal evolutionary patterns. We consider increased amounts of practical bioinformatics at school to be a priority.

Motivated by the increasing importance of bioinformatics to the life sciences and its appearance on school curricula, we conducted a preliminary investigation of the benefits of bringing university-level bioinformatics teaching material to voluntary groups of children in the last 2 years of school in Scotland (S5 and S6; pupils aged 15–17). The material was originally developed for an optional, final-year undergraduate module at the University of St Andrews, BL4273 Bioinformatics for Biologists (https://www.st-andrews.ac.uk/coursecatalogue/ug/2015-2016). To better match bioinformatics as it is actually used in research at universities, institutes and industry, the material uses the Linux operating system, in this case a variant of Rasbpian Linux running on low-cost Raspberry Pi hardware. This material has been released under an open access licence, as part of 4273π (Barker et al. 2013; http://4273pi.org). Our proposition was that school pupils can benefit from practical, undergraduate-level bioinformatics teaching material. Compared to the undergraduates for whom this material was originally developed, school pupils are less experienced and knowledgeable about biology in general. However, their levels of practical bioinformatics experience are broadly similar: zero in the case of the school pupils, and approximately ten actual contact hours among undergraduates at the time of starting the module.

Many of the skills developed in our activities, and 4273π or bioinformatics in general, are generic skills in computational science. For example, although the programming language taught in the "INTRO" component—Perl—is particularly widely used in bioinformatics (e.g. Stajich et al. 2002; Stabenau et al. 2004), it is structurally similar to other programming languages widely used in science, including C, Fortran, Java, Python and R. Use of the command-line, emphasised in 4273π, is also essential in computational physics, computational chemistry and, indeed, computer science. Although computational chemistry is not yet part of the Higher qualification in Chemistry, several simulations are suggested by the Scottish Qualifications Authority (2015a). Computational skills, as taught in 4273π, will be valuable to students taking chemistry, physics and other STEMM subjects at university.

Judged by pupil self-assessment forms, our preliminary trial was a success, though caution is required due to the small sample size. We will continue developing peer-reviewed bioinformatics material, targeted at school pupils and/or undergraduates, and applying it in practice. This will simultaneously lead to expansion of the 4273π resource and the gathering of larger, more complex and conclusive educational data at a future date. 4273π itself, and links to relevant social media groups, may be found at http://4273pi.org.

Methods

Two activities were carried out, each using a voluntary group of seven pupils studying science from a single school in Scotland. One group was from Kilgraston, an independent girls' school, and the other was from Forfar Academy, a comprehensive school. In the case of Kilgraston, five pupils were at S5 and two were at S6 level, and instruction and assistance were provided by D.B., M.M.C., G.T.P.M. and H.P. In the case of Forfar Academy, all pupils were at S5 level, of whom two where girls and five were boys, and instruction and assistance were provided by R.G.A., D.B., L.D., J.L.M. and S.D.S. Generally, university staff or PhD students provided detailed instruction on the bioinformatics activity, and school staff highlighted links to material already taught and the curriculum. With a combination of university staff or PhD students and school staff, students were guided through the practical material of two components (modules) of

Barker *et al. International Journal of STEM Education* (2015) 2:17

Page 3 of 8

4273π Bioinformatics for Biologists. At Kilgraston, the event was held at the school, occupying an entire day on which no other classes were scheduled. With Forfar Academy, the event was held at the University of St Andrews, where students participated in an afternoon and evening session, primarily held in the same room used, at other times, by undergraduates on the BL4273 module. Refreshment breaks were included, using the school's usual facilities (Kilgraston) or the Bell Pettigrew Museum (St Andrews; http://www.st-andrews.ac.uk/museum/bellpettigrew). In total, the teaching time was approximately 4 h.

Raspberry Pi Model B hardware was used, one per student (plus one connected to a projector for demonstration). Prior to the first event, at Kilgraston, tasks were selected from existing material in discussion between D.B. and H.P. (familiar with 4273π) and M.M.C. (familiar with the school curriculum). For both groups of pupils, the first task corresponded to the "INTRO" component of 4273π Bioinformatics for Biologists, providing an introduction to the Raspberry Pi computer hardware, the Linux command-line, BLAST sequence similarity search software and Programming in the Perl language. The second task corresponded to the "DNA" component, involving an introduction to the FlyBase database (Dos Santos et al. 2015) and genome annotation with BLAST (Altschul et al. 1997), GeneWise (Birney et al. 2004) and SNAP (Korf 2004). Hard-copy handouts were provided. The handouts for Kilgraston and Forfar Academy were identical in content apart from date, location of the event, staff details and location of files (~/kilgraston or ~/forfar_academy). For the record, the specific handouts used are available as Additional file 1 (Kilgraston) and Additional file 2 (Forfar Academy), but with names and contact details redacted. The latest, open access versions of these will be found in 4273π (http://4273pi.org).

Hard-copy, paired, "before" (prior to the use of the computers) and "after" questionnaires were used for pupils and school staff, involving questions on a 1–5 Likert scale for self-assessment of attitudes and free text (Table 1; Additional file 3). In preparing the questionnaires, for those questions on a Likert scale, the sequence of questions was randomised and the sense of each question ("1" corresponding to "good" on our subjective scale, vs "1" corresponding to "bad") was randomised. The same sequence and sense were used for each questionnaire handed out; within each group (pupils or staff), the sequence of paired questions was the same "before" and "after". Results of the questions on the Likert scale were summarised per question as a bar chart, and as a likelihood ratio sign test for evidence of systematic change over the course of the activity. We apply a likelihood approach to statistical inference (Birnbaum 1962; Edwards 1992; Royall 1997; Barker 2015). In common with other approaches to statistical inference, this provides no absolute

threshold beyond which evidence is considered conclusive. By convention, we define "strong" evidence as a log (ln) likelihood ratio, $\Delta\ell$, of at least 2, or a likelihood ratio of at least 8 (Edwards 1992, pp. 199–202; Royall 1997). Were $\Delta\ell$ converted to a p value under the assumptions of a likelihood ratio test (Wilks 1938), then for one free parameter $\Delta\ell \geq 2$ corresponds to $p \leq 0.046$, approximately the traditional threshold for statistical significance prior to any correction for multiple testing (i.e. $p < 0.05$). Calculations were performed in R (R Development Core Team 2010).

Results

Questionnaire responses, with paired before and after questionnaire answers, are available as Additional file 3 and summarised in Table 1. The spread of answers to pupil questionnaires, on the Likert scale, is presented in Fig. 1. Because individual pupils may apply different criteria for each specific question (i.e. categories are simultaneously pupil-specific and question-specific), we regard the separation into five categories as unsuitable for either a continuous approximation or ranking. However, the direction of change on the Likert scale is comparable throughout the data, on the weak assumption that each pupil applies his or her own criteria for a given question consistently both before and after. Apart from changes over the course of the activity, whether the weight of answers is in the "disagree" categories (1 and 2) vs the "agree" categories (4 and 5) may also have some general meaning.

To an extent, all our questions are expected, if anything, to improve as a result of the activities. For example, a significant part of the first component ("INTRO") is devoted to programming. Hence, it would represent either disaster for the educational approach, or perhaps humour on the part of pupils, if pupils tended to agree more with "I cannot program a computer" after the event than before. (The question has some similarity to a "control" in a laboratory experiment). Fortunately, in line with the educational approach being sound, evidence of disastrous effects was weak or absent.

With Question 1, "I will end up working in science", most pupils agree and there is no systematic change during the activity. This is as expected, because pupils with no interest in science would be unlikely to volunteer to take part. Question 2, "I think computers are useful within biology", is more specifically related to the activity itself. Again, most pupils agree both before and after, but here we see strong evidence of improvement (log likelihood ratio, $\Delta\ell = 3.47$). The usefulness of computers in other sciences (Question 9, "I think computers are useful within sciences other than biology") shows little change as a result of the activity. As part of the volunteers' positive mentality, we also see that most pupils expected to enjoy the activity and, in practice, did so (Question 4, "I do not expect to enjoy [after: did not enjoy] the activity today"). For

Barker *et al. International Journal of STEM Education* (2015) 2:17

Page 4 of 8

Table 1 Analysis of before and after questions for pupils on a Likert scale (1 "strongly disagree" to 5 "strongly agree"). For each pair of responses for each question, for each pupil, the change from before to after was noted, if any, and was reduced to a binary variable, indicating increase or decrease on the Likert scale. The proportion of changes that were increases and the proportion that were decreases constitute our maximum likelihood (ML) estimates for the probability of increase and probability of decrease, conditional on a change occurring. Whether the majority of changes are in the direction of improvement is indicated, with the direction indicating improvement being a subjective judgement by the authors in the context of this study. Assuming a binomial distribution, for each question separately, the likelihood of the observed changes was calculated, firstly, assuming the ML estimates for probability of increase and decrease obtained from the data; and secondly, assuming an extrinsic hypothesis that the probability of increase and probability of decrease are equal (0.5). From these likelihoods, the likelihood ratio and its natural logarithm ($\Delta\ell$) were calculated. Rows in italics show strong evidence of overall change in a specific direction over the course of the activity ($\Delta\ell \geq 2$). $N = 12$ pupils submitted both "before" and "after" questionnaires

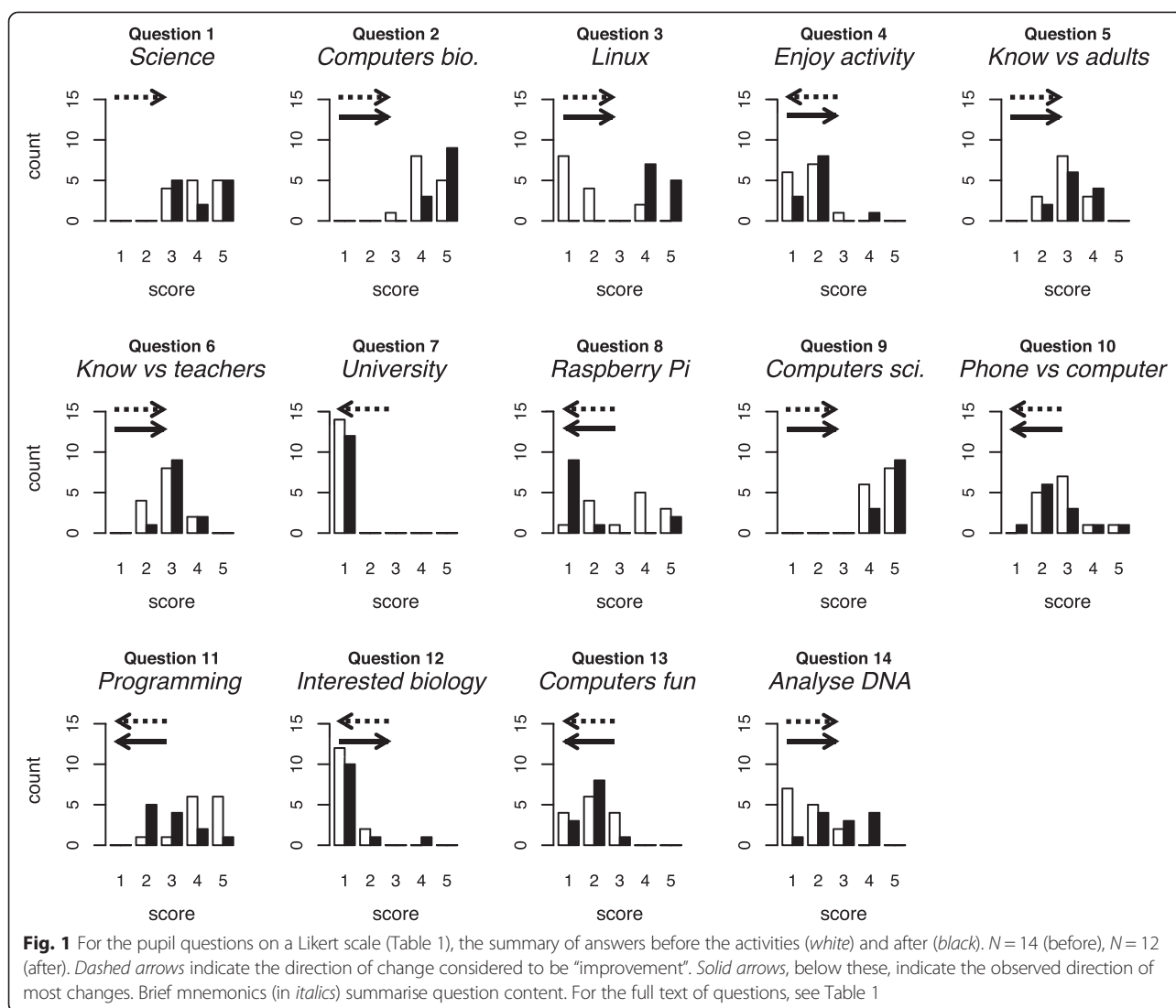| Question	Increases	Decreases	Increases proportion	Decreases proportion	Most changes were improvements?	Likelihood ratio	$\Delta\ell$
1. I will end up working in science.	1	1	0.5	0.5	No	1	0
2. I think computers are useful within biology.	*5*	*0*	*1*	*0*	*Yes*	*32*	*3.47*
3. I have heard of Linux.	*10*	*0*	*1*	*0*	*Yes*	*1024*	*6.93*
4. I do not expect to enjoy [after: did not enjoy] the activity today	4	3	0.57	0.43	No	1.07	0.07
5. I know more about computers than most adults do.	4	2	0.67	0.33	Yes	1.40	0.34
6. I know more about computers than my teachers do.	2	0	1	0	Yes	4	1.39
7. I am not intending to go to university.	0	0	n/a	n/a	n/a	1	0
8. I have not heard of the Raspberry Pi.	*1*	*10*	*0.09*	*0.91*	*Yes*	*71.78*	*4.27*
9. I think computers are useful within sciences other than biology.	2	0	1	0	Yes	4	1.39
10. I would rather not use a computer, I would prefer to use a phone or a tablet.	*0*	*3*	*0*	*1*	*Yes*	*8*	*2.08*
11. I cannot program a computer.	*1*	*8*	*0.11*	*0.89*	*Yes*	*22.17*	*3.10*
12. I am not interested in biology.	1	0	1	0	No	2	0.69
13. I do not enjoy using computers for fun.	1	2	0.33	0.67	Yes	1.19	0.17
14. I am good at using a computer to analyse DNA sequences.	*8*	*1*	*0.89*	*0.11*	*Yes*	*22.17*	*3.10*

n/a not available

this question, there is a fairly high level of change over the course of the activity, with four pupils reporting changes on the Likert scale in the direction opposite to "improvement", and three in the direction of "improvement" (Table 1). All pupils have a strong desire to continue studies at university, which was entirely unchanged by participating in the activity (Question 7, "I am not intending to go to university"). Both before and after, they have a strong interest in biology (Question 12, "I am not interested in biology"). One pupil reports a change in the direction opposite to improvement, whose implications are difficult to assess without a larger sample. All appear interested in computers (Question 13, "I do not enjoy using computers for fun").

Questions 3 ("I have heard of Linux") and 8 ("I have not heard of the Raspberry Pi") reveal that, before the activity, most pupils have not heard of the Linux operating system or the Raspberry Pi, though more have heard of the Raspberry Pi than Linux. Improvement over the course of the activity is almost guaranteed due to its content, and indeed there is strong evidence for this ($\Delta\ell = 6.93$ for Linux, $\Delta\ell = 4.27$ for the Raspberry Pi).

More complex and perhaps more educationally relevant (Question 11, "I cannot program a computer"), students show strong evidence of improvement from a position of not regarding themselves able to program a computer, to being positive or at least uncertain ($\Delta\ell = 3.10$). From a position of considering themselves generally poor at using a computer to analyse DNA sequences (Question 14, "I am good at using a computer to analyse DNA sequences"), there is strong evidence of improvement towards a more positive assessment of their own abilities ($\Delta\ell = 3.10$), although no pupil ended up strongly confident (Fig. 1). We speculate this would improve further with repeated activities.

Fig. 1 For the pupil questions on a Likert scale (Table 1), the summary of answers before the activities (*white*) and after (*black*). N = 14 (before), N = 12 (after). *Dashed arrows* indicate the direction of change considered to be "improvement". *Solid arrows*, below these, indicate the observed direction of most changes. Brief mnemonics (in *italics*) summarise question content. For the full text of questions, see Table 1

Questions 5 ("I know more about computers than most adults do") and 6 ("I know more about computers than my teachers do") are logically correlated, and both show weak evidence of improvement as a result of the activities. We prefer to interpret this as evidence of increased computational confidence among the pupils; however, heightened perception of adult incompetence cannot be ruled out.

Question 10 ("I would rather not use a computer, I would rather use a phone or a tablet") indicates a baseline level of openness to the educational method used. Our activity, in contrast to much computational activity by young people, is computer-based, using the Raspberry Pi. Pupils' views are generally moderate both before and after, but with borderline strong evidence of a strengthening tendency to prefer a computer as a result of the activity ($\Delta \ell = 2.08$).

Pupil free-text answers to Q15 ("What was the best part of the activity today, and why") included "getting the hands-on experience with computers, as it is not something we ever get to do at school"; "I liked the programming—making files and commands—because it was really interesting to see how computers actually work"; "The simpler programming stages as they were at an accomplishable level & well explained"; "Learning how computers are used in biology and getting to try it ourselves because it was interesting and a new experience."; "The best part was going on flybase because I found it really interesting". Free-text answers to Q16 ("What was the worst part of the activity today, and why?") included "Typing everything during programme because it took ages and a tiny mistake or omission could cause problems."; "The worst part was flybase as it was difficult to understand."; "I enjoyed the entire day and there wasn't

Barker *et al. International Journal of STEM Education* (2015) 2:17

Page 6 of 8

one part I didn't enjoy. It was really interesting."; "typing in all of the commands to the LXterminal as it took a long time and it was easy to make mistakes."; "I didn't enjoy typing stuff into LxTerminal because I was awful at it"; "The DNA sequencing on the internet was confusing and poorly explained in the handbook"; and "I don't think the instructions were very clear" (Additional file 3).

It seems the activities stretched pupils intellectually and practically, which was regarded both positively and negatively among the pupils. This is illustrated, for example, by opposite views on the Flybase database, and may also be reflected in changing views on the enjoyability of the activities (Question 4). Typing programs and using the command-line within the terminal was unpopular. However, this is a crucial part of computational science, which cannot be bypassed, and which remains difficult even for professional bioinformaticians. The practical handouts clearly did not suit all pupils. These attempt a practical balance between brevity and total coverage. In future, these could be improved by inclusion of diagrams and screenshots.

School staff free-text answers to the question "What was the best part of the activity today, and why?" included "The pupils being given the opportunity to contextualise their theory based work in class into a practical application using the software. They gained a great deal of practice in making the connection between proteins + amino acids being the result of DNA sequences." And: "Watching pupils who had never written code program Pis. Seeing their excitement when the code worked." Also: "The pupils were all fully engaged & gained a great deal from the seminar. I have also gained an insight into bioinformatics which will be vital for teaching the next part of the new higher on genomics" (Additional file 3).

Discussion

Our results were for a small sample and cannot represent the full range of voluntary groups one could draw from schools even in Scotland. But it is encouraging for the approach taken that almost all changes in pupil questionnaires, over the course of the activity, were in the direction that we regard as indicating improvement.

For all pupil questions showing strong evidence of change over the course of the activity ($\Delta\ell \geq 2$), the tendency was in the direction of improvement. Also, free-text responses from pupils were generally positive, though with a suggestion that the handouts could be improved ("Results", above). On this basis, we judge the preliminary investigation a success. The evidence strongly suggests that, in their own assessment, the pupils benefited from the bioinformatics activities undertaken. This complements other research, involving successful bioinformatics activities developed specifically for school level (e.g. Marques et al. 2014).

Gallagher et al. (2011) also reported a successful experience bringing bioinformatics to high school level, but with three challenges. We did not experience any of these. Firstly, some of the students in their study questioned the relevance of computation to biology. Gallagher et al. propose that part of their difficulty was use of computer scientists to teach lessons. In our case, lessons were co-taught by university biologists or biologists and chemists, alongside school staff, lending support to this proposition. However, other factors may be relevant, for example the use of whole-class teaching by Gallagher et al., as opposed to our voluntary groups; and the passage of time since their study. Secondly, some of the students in the study by Gallagher et al. doubted the value of learning about algorithms. Since our material was practical, with little detail on algorithms, the objection would—a priori—be unlikely. Thirdly, Gallagher et al. found some students doubted the relevance of bioinformatics material if this was not covered in official tests. In our case, there is relevance to the syllabus (see "Relevance to the new Higher curriculum for Biology and Human Biology", below), and this was highlighted during the activities by school staff. With a different audience but similar educational aim, Wood and Gebhardt (2013) also emphasise the importance of relevance to the curriculum. They report successfully updating high school teachers on the topic of bioinformatics via hands-on experience, practical demonstrations and lectures. One of their considerations for those wishing to replicate their approach, "Work within the practical limitations of the classroom", may be alleviated by hardware such as the Raspberry Pi (see "The Raspberry Pi—a flexible, general-purpose computer", below).

How far can we generalise from our results? Further work is warranted, for example using a broader range of teaching material; involvement of a larger number of voluntary groups from a wider range of schools, for a more varied sample and greater statistical power; use of less enthusiastic staff; a lower staff/student ratio; a whole-class activity, not only a voluntary class sample; validation of the assessment instrument; direct measurement of student knowledge acquisition, in addition to self-assessment; interviews with staff and students and field observations. Where there is a surfeit of material available among our university colleagues—for example, introductory teaching material on the widely used program BLAST (Altschul et al. 1997)—one can imagine educational experiments, allowing an evidence-based decision on which material is more successful, perhaps in different geographical regions, countries, educational systems or types of school or with pupils of different age, academic achievement levels, social background or gender. These are important topics for the future.

Barker *et al. International Journal of STEM Education* (2015) 2:17

Page 7 of 8

Bioinformatics has always been relatively "open" as a subject, with public deposition of DNA sequence data associated with research publications and a wealth of free software. Although the algorithms do not change rapidly, the precise details of the software do. For these two reasons—the tradition of openness within the field and the rapid obsolescence of specific details—we consider the subject particularly amenable to online-only, open access teaching material. This costs nothing to obtain; for those writing the material, it can be relatively easily updated in light of developments in software and data. The material we used was from one open access bioinformatics educational effort, 4273π (Barker et al. 2013). We hope to expand this resource to cover a wider range of bioinformatics topics, whilst continuing to ensure that all material is tested in practice and peer-reviewed.

Relevance to the new Higher curriculum for Biology and Human Biology

The new Scottish Qualifications Authority Curriculum for Excellence Higher Syllabus for both Higher Biology and Higher Human Biology includes DNA technologies. Part of the content focuses on the use of computer software in both Biology (Scottish Qualifications Authority 2014a, pp. 10, 16) and Human Biology (Scottish Qualifications Authority 2014b), pp. 14, 15, 56). The suggested learning activity "Use genome data to identify stop and start codons and known protein coding sequences" (Scottish Qualifications Authority 2014b) is directly addressed by the "DNA" component of 4273π, as used in the current study. Other components are also relevant, for example to phylogeny and comparative genomics in Higher Biology (Scottish Qualifications Authority 2014a, p. 16). In addition, the new Curriculum for Excellence Advanced Higher Biology has content related to protein structure (covered, to an extent, by the "ENZYME" component of 4273π) and scientific investigative approaches (Scottish Qualifications Authority 2015b).

Based on our experiences as outlined in this paper, and in agreement with Gallagher et al. (2011), workshops taught alongside university staff seem an efficient means of professional development for teachers, who may be presented with curriculum content which is new to them and outwith their experience. From the pupils' point of view, having access to workshops and seminars run by university staff also gives some experience of the university experience, particularly when—as in the current case—the teaching material is precisely at the undergraduate level.

The Raspberry Pi—a flexible, general-purpose computer

The use of Raspberry Pi computers allows for a range of new software and experiences for the pupils, without any major impact on school's ICT staff and systems. By bringing in complete software plus hardware systems, there is no need for specific software to be installed on school computers (some of which may require existing security restrictions to be relaxed). As the first part of the activity was to assemble the Pi, for the activity held at Kilgraston, the only prerequisite for the school was to check the school's VGA monitors would work with the Pi (http://4273pi.org/files/2015/10/hardware_4273pi.pdf) and check there were enough power sockets.

Pupils may be more confident using the Raspberry Pi, with the knowledge that if they "broke" the software set-up, instructors could just swap its SD card. This allows a risk-taking approach to learning, and the possibility to learn by ones mistakes in a way that is prevented in managed computer classrooms. (Despite this, in the two school activities in this study, and in 3 years of teaching BL4273 Bioinformatics for Biologists at the University of St Andrews, we have never had to reinitialise an SD card due to such an accident. We plan to incorporate greater software risks in future material). It also gives pupils experience of Linux, the operating system they may well use should they choose to further their career in computational science.

Conclusion

In our current study, we have demonstrated that practical bioinformatics material initially developed for an optional, final-year undergraduate module can be used successfully with school pupils at sixth form level (S5 and S6 in Scotland). Covering practical bioinformatics at this level in all schools may be an attainable goal, with subsequent benefits to higher education and research, in both biology and other sciences.

Additional files

> **Additional file 1: Teaching material and blank questionnaires used for the event with Kilgraston School.** Names and contact details have been redacted. (ZIP 432 kb)
>
> **Additional file 2: Teaching material and blank questionnaires used for the event with Forfar Academy.** Names and contact details have been redacted. (ZIP 478 kb)
>
> **Additional file 3: Data from questionnaires (LibreOffice Calc spreadsheet).** For details of contents, see the "Notes" worksheet. (ODS 82 kb)

Abbreviations

BLAST: Basic Local Alignment Search Tool; GOBLET: Global Organisation for Bioinformatics Learning, Education and Training; ICT: information and communications technology; ML: maximum likelihood; SD: secure digital; STEMM: Science, Technology, Engineering, Mathematics and Medicine; UK: United Kingdom; USA: United States of America.

Competing interests

The authors declare that they have no competing interests.

Barker *et al. International Journal of STEM Education* (2015) 2:17

Page 8 of 8

Author details
[1]Sir Harold Mitchell Building, School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, UK. [2]Biomedical Sciences Research Complex and EaStCHEM School of Chemistry, Purdie Building, Universityof St Andrews, North Haugh, St Andrews, Fife KY16 9ST, UK. [3]Current address: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. [4]Kilgraston School, Bridge of Earn, Perthshire PH2 9BQ, UK. [5]Current address: Forfar Academy, Taylor Street, Forfar DD8 3LB, UK. [6]Forfar Academy, Taylor Street, Forfar DD8 3LB, UK. [7]Current address: Portlethen Academy, Cookston Road, Portlethen, Aberdeenshire AB12 4PT, UK.

References
Altschul, SF, Madden, TL, Schäffer, AA, Zhang, J, Zhang, Z, Miller, W, & Lipman, DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research, 25*, 3389–3402.

Barker, D. (2015). Seeing the wood for the trees: philosophical aspects of classical, Bayesian and likelihood approaches in statistical inference and some implications for phylogenetic analysis. *Biology & Philosophy, 30*, 505–525.

Barker, D, Ferrier, DEK, Holland, PWH, Mitchell, JBO, Plaisier, H, Ritchie, MG, & Smart, SD. (2013). 4273π: bioinformatics education on low-cost ARM hardware. *BMC Bioinformatics, 14*, 243.

Benson, DA, Clark, K, Karsch-Mizrachi, I, Lipman, DJ, Ostell, J, & Sayers, EW. (2015). GenBank. *Nucleic Acids Research, 43*, D30–D35.

Birnbaum, A. (1962). On the foundations of statistical inference. *Journal of the American Statistical Association, 57*, 269–306.

Birney, E, Clamp, M, & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research, 14*, 988–995.

College voor Examens (2014). Biologie VWO. Syllabus Centraal Examen 2016. http://www.examenblad.nl/examenstof/syllabus-2016-biologie-vwo-nader/2016/vwo/f=/biologie_vwo_2016_def_voor_hervaststelling.pdf. . Accessed 21 October 2015.

Corpas, M, Jimenez, RC, Bongcam-Rudloff, E, Budd, A, Brazas, MD, Fernandes, PL, Gaeta, B, van Gelder, C, Korpelainen, E, Lewitter, F, McGrath, A, MacLean, D, Palagi, PM, Rother, K, Taylor, J, Via, A, Watson, M, Schneider, MV, & Attwood, TK. (2015). The GOBLET training portal: a global repository of bioinformatics training materials, courses and trainers. *Bioinformatics, 31*, 140–142.

Cunningham, F, Amode, MR, Barrell, D, Beal, K, Billis, K, Brent, S, Carvalho-Silva, D, Clapham, P, Coates, G, Fitzgerald, S, Gil, L, García Girón, C, Gordon, L, Hourlier, T, Hunt, SE, Janacek, SH, Johnson, N, Juettemann, T, Kähäri, AK, Keenan, S, Martin, FJ, Maurel, T, McLaren, W, Murphy, DN, Nag, R, Overduin, B, Parker, A, Patricio, M, Perry, E, Pignatelli, M, Singh Riat, H, Sheppard, D, Taylor, K, Thormann, A, Vullo, A, Wilder, SP, Zadissa, A, Aken, BL, Birney, E, Harrow, J, Kinsella, R, Muffato, M, Ruffier, M, Searle, SMJ, Spudich, G, Trevanion, SJ, Yates, A, Zerbino, DR, & Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research, 43*, D662–D669.

Dos Santos, G, Schroeder, AJ, Goodman, JL, Strelets, VB, Crosby, MA, Thurmond, J, Emmert, DB, Gelbart, WM, & The FlyBase Consortium (2015). FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Research, 43*, D690-D697

Edwards, AWF. (1992). *Likelihood* (expanded ed.). Baltimore: Johns Hopkins University Press.

Gallagher, SR, Coon, W, Donley, K, Scott, A, & Goldberg, DS. (2011). A first attempt to bring computational biology into advanced high school biology classrooms. *PLoS Computational Biology, 7*, e1002244.

Galperin, MY, Rigden, DJ, & Fernández-Suárez, XM. (2015). The 2015 *Nucleic Acids Research* database issue and molecular biology database collection. *Nucleic Acids Research, 43*, D1–D5.

Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. *PLoS Computational Biology, 7*, e1002021.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics, 5*, 59.

Kovarik, DN, Patterson, DG, Cohen, C, Sanders, EA, Peterson, KA, Porter, SG, & Chowning, JT. (2013). Bioinformatics education in high school: implications for promoting science, technology, engineering, and mathematics careers. *CBE – Life Sciences Education, 12*, 441–459.

Lewitter, F, & Bourne, PE. (2011). Teaching bioinformatics at the secondary school level. *PLoS Computational Biology, 7*, e1002242.

Machluf, Y, & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics, 14*, 648–660.

Magana, AJ, Taleyarkhan, M, Rivera Alvarado, D, Kane, M, Springer, J, & Clase, K. (2014). A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research. *CBE – Life Sciences Education, 13*, 607–623.

Marques, I, Almeida, P, Alves, R, João Dias, M, Godinho, A, & Pereira-Leal, JB. (2014). Bioinformatics projects supporting life-sciences learning in high schools. *PLoS Computational Biology, 10*, e10hardware_4273pi.pdf404.

McQueen, J, Wright, JJ, & Fox, JA. (2012). Design and implementation of a genomics field trip program aimed at secondary school students. *PLoS Computational Biology, 8*, e1002636.

R Development Core Team (2010). R: A language and Environment for Statistical Computing. http://www.R-project.org.

Royall, R. (1997). *Statistical evidence: a likelihood paradigm*. Boca Raton: Chapman & Hall/CRC.

Scottish Qualifications Authority (2014a). Higher Biology Course Support Notes. http://www.sqa.org.uk/sqa/48674.html?subject=Biology. Accessed . 20 October 2015.

Scottish Qualifications Authority (2014b). Higher Human Biology Course Support Notes. http://www.sqa.org.uk/sqa/48674.html?subject=Human%20Biology . Accessed 20 October 2015.

Scottish Qualifications Authority (2015a). Higher Chemistry Course Support Notes. http://www.sqa.org.uk/files_ccc/CfE_CourseUnitSupportNotes_Higher_Sciences_Chemistry.pdf. Accessed 20 October 2015.

Scottish Qualifications Authority (2015b). Advanced Higher Biology Course/Unit Support Notes. http://www.sqa.org.uk/sqa/48674.html?subject=Biology. Accessed 20 October 2015.

Stabenau A, McVicker G, Melsopp C, Proctor G, Clamp M, & Birney E. (2004). The Ensembl core software libraries. *Genome Research, 14*, 929–933.

Stajich, JE, Block, D, Boulez, K, Brenner, SE, Chervitz, SA, Dagdigian, C, Fuellen, G, Gilbert, JGR, Korf, I, Lapp, H, Lehväslaiho, H, Matsalla, C, Mungall, CJ, Osborne, BI, Pocock, MR, Schattner, P, Senger, M, Stein, LD, Stupka, E, Wilkinson, MD, & Birney, E. (2002). The BioPerl toolkit: Perl modules for the life sciences. *Genome Research, 12*, 1611–1618.

Toby, I, & Pope, A. (2014). Bioinformatics tools available for K-12 students to engage in research. *Aviation, Space and Environmental Medicine, 85*, 484–485.

Wefer, SH, & Sheppard, K. (2008). Bioinformatics in high school biology curricula: a study of state science standards. *CBE – Life Sciences Education, 7*, 155–162.

Wilks, SS. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9*, 60–62.

Wood, L, & Gebhardt, P. (2013). Bioinformatics goes to school – new avenues for teaching contemporary biology. *PLoS Computational Biology, 9*, e10hardware_4273pi.pdf089.